

GenCore version 5.1.3
Copyright (c) 1993 - 2003 Compen Ltd.

OM protein - protein search, using sw model

Run on: January 3, 2003, 13:01:31 ; Search time 13.3043 Seconds
(without alignments)
112.230 Million cell updates/sec

Title: US-09-801-784A-1

Perfect score: 173
Sequence: 1 VEKNITVTASVDPITDLDGSLPSAVALTSPSA 36

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Number of hits satisfying chosen parameters: 112892

Total number of hits satisfying chosen parameters: 112892

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SwissProt_40.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	155	89.6	170	1 FMC1_ECOLI	P02371 escherichia
2	153	88.4	171	1 FMS1_ECOLI	P25730 escherichia
3	60.5	35.0	714	1 YFET_YEAST	P43556 saccharomyc
4	58	33.5	1051	1 PTK7_CHICK	Q91048 gallus gall
5	56.5	32.7	580	1 GGT_ECOLI	P19956 escherichia
6	53	30.6	360	1 VTPX_TTVI	P19274 thermoprote
7	53	30.6	474	1 VTP3_TTVI	P19275 thermoprote
8	52.5	30.3	243	1 YX42_MYCTU	O53392 mycobacteri
9	52	30.1	514	1 CSA_DICDI	P08796 dictyostell
10	52	30.1	773	1 YHGF_ECOLI	P46837 escherichia
11	52	30.1	952	1 UVRA_THETH	O56242 thermus the
12	50	28.9	438	1 YDR6_SCHPO	O13741 schizosacch
13	50	28.9	481	1 PYR5_MOUSE	P13439 mus musculu
14	50	28.9	848	1 YAM6_SCHPO	Q10061 schizosacch
15	49	28.3	925	1 W70T_HUMAN	P57737 homo sapien
16	49	28.3	281	1 H1S2_ARATH	O82768 arabidopsis
17	49	28.3	299	1 YJH9_YEAST	P47032 saccharomyc
18	49	28.3	384	1 UDG_STRPN	O57346 streptococc
19	49	28.3	466	1 EXU_DROVI	O24747 dirosophila
20	49	28.3	1070	1 PTK7_HUMAN	Q13308 homo sapien
21	48.5	28.0	2363	1 SPCO_MOUSE	O62651 mus musculu
22	48	27.7	172	1 YH74_METH	O27862 methanobact
23	48	27.7	299	1 MRAB_MYCPU	O98675 mycoplasma
24	48	27.7	326	1 MTC2_CHVPI	P31118 parametium
25	48	27.7	349	1 RECA_AZOVI	P29246 azobacter
26	48	27.7	1744	1 CO4_HUMAN	P01028 homo sapien
27	48	27.7	1888	1 CALE_CHICK	P32018 gallus gall
28	48	27.7	2491	1 MPRI_HUMAN	P11717 homo sapien
29	47.5	27.5	1130	1 ITA6_HUMAN	P23329 homo sapien
30	47	27.2	189	1 TBP_PYROG	P58127 pyrodictium
31	47	27.2	390	1 UPXB_HAEIN	P45011 haemophilus
32	47	27.2	480	1 PYR5_HUMAN	P11172 homo sapien
33	47	27.2	1102	1 TRI3_YEAST	Q03660 saccharomyc

ALIGNMENTS

RSRUT 1					
FMCI_ECOLI		STANDARD:	PRT:	170 AA.	
AC P02971:					
DT 21-JUN-1986 (Rel. 01, Created)					
DT 01-FEB-1991 (Rel. 17, Last sequence update)					
DT 15-JUL-1999 (Rel. 38, Last annotation update)					
DE CFA/I fimbrial subunit B precursor (Colonization factor antigen 1					
DE subunit B) (CFA/I pili) (CFA/I antigen).					
GN CFA8.					
OS Escherichia coli.					
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;					
OC Escherichia.					
OX NCBI_TaxID=562;					
RN [1]					
RP SEQUENCE FROM N.A.					
RX MEDLINE=89173309; PubMed=2564374;					
RA Karjalainen T.K., Evans D.G., So M., Lee C.-H.;					
RT "Molecular cloning and nucleotide sequence of the colonization factor					
RT antigen I gene of Escherichia coli.";					
RL Infect. Immun. 57:1126-1130(1989).					
RN [2]					
RP SEQUENCE FROM N.A.					
RX MEDLINE=89330163; PubMed=2569152;					
RA Hamers A.M., Pel H.J., Willshaw G.A., Kusters J.G.;					
RA van der Zeijst B.A.M., Gaastera M.;					
RT "The nucleotide sequence of the first two genes of the CFA/I fimbrial					
RT operon of human enterotoxigenic Escherichia coli.";					
RL Microb. Pathog. 6:297-309(1989).					
RN [3]					
RP SEQUENCE OF 24-170 FROM N.A.					
RX MEDLINE=82235736; PubMed=6124420;					
RA STRAIN=O78:H11 / H10407;					
RX MEDLINE=90036735; PubMed=2572583;					
RA Hall R.H., Maneval D.R. Jr., Collins J.H., Theibert J.L.;					
RA Levine M.M.;					
RT "Purification and analysis of colonization factor antigen 1, coli					
RT surface antigen 1, and coli surface antigen 3 fimbriae from					
RT enterotoxigenic Escherichia coli.";					
RL J. Bacteriol. 171:6372-6374(1989).					
RL [4]					
RP SEQUENCE OF 24-42.					
RX STRAIN=O78:H11 / H10407;					
RX MEDLINE=90036735; PubMed=2572583;					
RA Hall R.H., Maneval D.R. Jr., Collins J.H., Theibert J.L.;					
RA Levine M.M.;					
RT "Purification and analysis of colonization factor antigen 1, coli					
RT surface antigen 1, and coli surface antigen 3 fimbriae from					
RT enterotoxigenic Escherichia coli.";					
RL J. Bacteriol. 171:6372-6374(1989).					
RL [5]					
RP FUNCTION: FIMBRIAE (ALSO CALLED PILI), POLAR FILAMENTS RADIATING					
RP FROM THE SURFACE OF THE BACTERIUM TO A LENGTH OF 0.5-1.5					
RP MICROMETERS AND NUMBERING 100-300 PER CELL, EVABLE BACTERIA TO					
RP COLONIZE THE EPITHELIUM OF SPECIFIC HOST ORGANS.					
RP SUBUNIT: CFA/I FIMBRIAE ARE RATHER RIGID, THREAD-LIKE FILAMENTS OF					
RP 0.5-1 MICROMETER, WITH AN APPARENT AXIAL HOLE, AND A DIAMETER OF					
RP 7 NANOMETERS. A SINGLE CFA/I FIMBRIA CONSISTS OF ABOUT 100					
RP IDENTICAL PROTEIN SUBUNITS.					

Q9PJY2 chiomydia m
P16271 lactococcus
P15293 lactococcus
P14912 lactococcus
P14913 petroselinu
P08155 dirosophila
Q06278 homo sapien
P3869 synechocyst
P4256 aspergillus
P32951 candida par
P31120 escherichia

```
CC -!- INDUCTION: CFA/II FIMBRIAE ARE ONLY EXPRESSED IN THE PRESENCE OF
CC THE POSITIVE REGULATOR CFAD.
CC -!- SIMILARITY: TO THE CSI FIMBRIAL SUBUNIT A (CSOA).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; M55661; AAC41415.1; -.
DR PIR; A30589; YQECCL.
KW Fimbria; Antigen; Plasmid; Signal.
FT SIGNAL 1 23
FT CHAIN 24 170 CFA/II FIMBRIAL SUBUNIT B.
FT CONFLICT 37 37 V -> A (IN REF. 2).
FT CONFLICT 76 76 D -> N (IN REF. 3).
FT CONFLICT 97 97 S -> A (IN REF. 3).
SQ SEQUENCE 170 AA; 17461 MW; 924347F07F33CDAB CRC64;

Query Match 89.6%; Score 155; DB 1; Length 170;
Best Local Similarity 88.9%; Pred. No. 2.2e-14;
Matches 32; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

QY 1 VEKNITVTASVDPTIDLLQAGSALPSAVALTYSPA 36
| | | | | | | | | | | | | | | | | | | | | |
Db 24 VEKNITVTASVDPTIDLLQAGSALPSAVALTYSPA 59

RESULT 2
FMS1_ECOLI
ID FMS1_ECOLI STANDARD; PRT; 171 AA.
AC P25730.
DT 01-MAY-1992 (Rel. 22, Created)
DT 01-MAY-1992 (Rel. 22, Last sequence update)
DT 01-OCT-1996 (Rel. 34, Last annotation update)
DE CSI fimbrial subunit A precursor (CSI pilin).
CN CSOA OR COOA.
OS Escherichia coli.
OC Plasmid pDEP23, and Plasmid pEU605.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Escherichia.
OX NCBI_TaxID=562;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN:O139:H28; PLASMID=pDEP23;
RX MEDLINE=91157443; PubMed=1679404;
RA van der Zeijst B.A.M., van Vliet A.H.M., Willshaw G.A.,
RA "Analysis of the first two genes of the CSI fimbrial operon in human
RT enterotoxigenic Escherichia coli of serotype O139:H28";
RL FEMS Microbiol. Lett. 64:265-270(1991).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN:O6:H16; PLASMID=pEU605;
RX MEDLINE=91034170; PubMed=1977705;
RA Perez-Casal J., Swartley J.S., Scott J.R.;
RA "Gene encoding the major subunit of CSI pilin of human enterotoxigenic
RT Escherichia coli.";
RL Infect. Immun. 59:3594-3600(1990).
RN [3]
RP SEQUENCE OF 24-42.
RC STRAIN=60R75;
RX MEDLINE=90036735; PubMed=2572583;
RA Hall R.H., Maneval D.R. Jr., Collins J.H., Theibert J.L.,
RA Levine M.M.;
RA "Purification and analysis of colonization factor antigen 1, coli
RT surface antigen 1, and coli surface antigen 3 fimbriae from
RT enterotoxigenic Escherichia coli.";
RL J. Bacteriol. 171:6372-6374(1989).
CC -!- FUNCTION: FIMBRIAE (ALSO CALLED PILI), POLAR FILAMENTS RADIATING
```

```
CC FROM THE SURFACE OF THE BACTERIUM TO A LENGTH OF 0.5-1.5
CC MICROMETERS AND NUMBERING 100-300 PER CELL, ENABLE BACTERIA TO
CC COLONIZE THE EPITHELIUM OF SPECIFIC HOST ORGANS.
CC -!- INDUCTION: CSI FIMBRIAE ARE ONLY EXPRESSED IN THE PRESENCE OF THE
CC POSITIVE REGULATOR RNS.
CC -!- SIMILARITY: TO THE CFA/II FIMBRIAL SUBUNIT B (CFAB).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; X62879; CAA44673.1; -.
DR EMBL; M58550; AAA23596.1; -.
DR PIR; A41467; A41467.
DR PIR; S19003; S19003.
KW Fimbria; Signal; Plasmid.
FT SIGNAL 1 23
FT CHAIN 24 171 CSI FIMBRIAL SUBUNIT A.
FT CONFLICT 28 28 I -> E (IN REF. 3).
SQ SEQUENCE 171 AA; 17542 MW; 46E70BE7053DBE13 CRC64;

Query Match 88.4%; Score 153; DB 1; Length 171;
Best Local Similarity 83.3%; Pred. No. 4.2e-14;
Matches 30; Conservative 5; Mismatches 1; Indels 0; Gaps 0;

QY 1 VEKNITVTASVDPTIDLLQAGSALPSAVALTYSPA 36
| | | | | | | | | | | | | | | | | | | | | |
Db 24 VEKNITVTASVDPTIDLLQAGSALPSAVALTYSPA 59

RESULT 3
YFE7_YEAST
ID YFE7_YEAST STANDARD; PRT; 714 AA.
AC P43556;
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Hypothetical 82.2 kDa protein in EMP47-SEC53 intergenic region.
CN YFL047W.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
OX NCBI_TaxID=4932;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=S288c / AB972;
RX MEDLINE=95400292; PubMed=7670463;
RA Murakami Y., Naitou M., Hagiwara H., Shibata T., Ozawa M.,
RA Sasanuma S.-I., Sasanuma M., Tsuchiya Y., Soeda E., Yokoyama K.,
RA Yamazaki M., Tashiro H., Eki T.;
RT "Analysis of the nucleotide sequence of chromosome VI from
RT Saccharomycetes cerevisiae.";
RL Nat. Genet. 10:261-268(1995).
CC -!- SIMILARITY: CONTAINS 1 DEP DOMAIN.
CC -!- SIMILARITY: CONTAINS 1 PCH DOMAIN.
CC -!- SIMILARITY: CONTAINS 1 RHO-GAP DOMAIN.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; D50617; BAA09194.1; -.
DR SGD; S0001847; YFL047W.
DR InterPro; IPR001060; Cdc15_Fes_C1P4.
DR InterPro; IPR000591; DEP.
```


RA Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,
RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,
RA Mau B., Shao Y.;
RT "The complete genome sequence of *Escherichia coli* K-12.";
RL Science 277:1453-1474(1997).
RN [3]
RP SEQUENCE OF 1-362 FROM N.A.
RC STRAIN-K12 / MG1655;
RX MEDLINE=94316500; PubMed=8041620;
RA Sofia H.J., Burland V., Daniels D.L., Plunkett G. III, Blattner F.R.;
RT "Analysis of the *Escherichia coli* genome. V. DNA sequence of the
RT region from 76.0 to 81.5 minutes.";
RL Nucleic Acids Res. 22:2576-2586(1994).
RN [4]
RP MUTAGENESIS.
RX MEDLINE=91080553; PubMed=1360205;
RA Hashimoto W., Suzuki H., Nohara S., Kumagai H.;
RT "Escherichia coli gamma-glutamyltranspeptidase mutants deficient in
RT processing to subunits.";
RL Biochem. Biophys. Res. Commun. 189:173-178(1992).
CC -1- CATALYTIC ACTIVITY: (5-L-glutamyl)-peptide + an amino acid =
CC peptide + 5-L-glutamyl-amino acid.
CC -1- PATHWAY: GGT PLAYS A KEY ROLE IN THE GAMMA-GLUTAMYL CYCLE, A
CC PATHWAY FOR THE SYNTHESIS AND DEGRADATION OF GLUTATHIONE, WHICH ARE
CC SYNTHESIZED IN PRECURSOR FORM FROM A SINGLE POLYPEPTIDE.
CC -1- SUBCELLULAR LOCATION: Periplasmic.
CC -1- SIMILARITY: STRONG, TO OTHER GGT ENZYMES AND TO GL-7ACA ACYLASES.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; M28722; AAA23869.1; -;
CC EMBL; U18997; AAA58245.1; -;
CC EMBL; AE000421; AAC76472.1; -;
CC EMBL; U00039; AAB18422.1; -;
CC PIR; J00028; EKECEX.
CC MEROPS; T03.061; -;
CC EcoGene; EG10374; ggt.
CC InterPro; IPR000101; Gglutnrsptdase.
CC Pfam; PF01019; G_glu_transpept; 1.
CC PRINTS; PR01210; GGTTRANSPTASE.
CC TIGRPFAMs; TIGR00066; g_glu_trans; 1.
DR PROSITE; PS00462; G_GLU_TRANSPEPTIDASE; 1.
KW Transferase; Acyltransferase; Periplasmic; Zymogen; Signal;
KW Glutathione biosynthesis; Complete proteome.
FT SIGNAL 1 25
FT CHAIN 26 390 GAMMA-GLUTAMYLTRANSFERASE LARGE CHAIN.
FT CHAIN 391 580 GAMMA-GLUTAMYLTRANSFERASE SMALL CHAIN.
FT BINDING 465 465 GAMMA-GLUTAMYL (POTENTIAL).
FT MUTAGEN 513 513 R->A: NOT PROCESSED INTO ITS SUBUNITS.
FT MUTAGEN 571 571 R->G: NOT PROCESSED INTO ITS SUBUNITS.
SQ SEQUENCE 580 AA; 61768 MW; 772F652EBA2A5F00 CRC64;
Query Match 32.7%; Score 56.5; DB 1; Length 580;
Best Local Similarity 40.0%; Pred. No. 4;
Matches 14; Conservative 9; Mismatches 7; Indels 5; Gaps 1;
OY 7 VTASVDPDT-----IDLQAGSALPSAVALTYSPA 36
DB 50 MWASVDATQGVVDILKEGGNAVDAAVAGYALA 84
RESULT 6
ID VTP3 TTV1 STANDARD; PRT; 360 AA.
AC P19274;
DT 01-NOV-1990 (Rel. 16, Created)

DT 01-NOV-1990 (Rel. 16, Last sequence update)
DT 01-FEB-1991 (Rel. 17, Last annotation update)
DE Viral protein TPX.
OS Thermoproteus tenax virus 1 (strain KRA1) (TTV1).
OC Viruses; dsDNA viruses, no RNA stage; Lipothirixviridae;
OC Lipothirixvirus.
OX NCBI_TaxID=10480;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=90174928; PubMed=2308830;
RA Neumann H., Zillig W.;
RT "The TTV1-encoded viral protein TPX: primary structure of the gene
RT and the protein.";
RL Nucleic Acids Res. 18:195-195(1990).
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; X14855; CAA33002.1; -;
CC PIR; S12850; S12850.
FT DOMAIN 278 353 3 THR-PRO REPEATS REGIONS AND TWO NEAR
FT IDENTICAL REPEATS.
FT REPEAT 270 291 THR-PRO(N).
FT REPEAT 292 301 THR-PRO(N).
FT REPEAT 302 322 THR-PRO(N).
FT REPEAT 323 332 THR-PRO(N).
FT REPEAT 333 353 THR-PRO(N).
SQ SEQUENCE 360 AA; 38214 MW; EF54FBCB753ED1CA CRC64;
Query Match 30.6%; Score 53; DB 1; Length 360;
Best Local Similarity 42.1%; Pred. No. 6.9;
Matches 16; Conservative 7; Mismatches 9; Indels 6; Gaps 2;
OY 3 KNITVTAS-----VDPTIDLLQAGSALPSAVALTYSP 35
DB 63 QSITITASSCTFNIDFTIALYNGSSY-SNLTLCVSP 99
RESULT 7
ID VTP3 TTV1V STANDARD; PRT; 474 AA.
AC P19275;
DT 01-NOV-1990 (Rel. 16, Created)
DT 01-NOV-1990 (Rel. 16, Last sequence update)
DT 01-FEB-1991 (Rel. 17, Last annotation update)
DE Viral protein TPX.
OS Thermoproteus tenax virus 1 (strain VT3) (TTV1).
OC Viruses; dsDNA viruses, no RNA stage; Lipothirixviridae;
OC Lipothirixvirus.
OX NCBI_TaxID=10481;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=90245666; PubMed=2336394;
RA Neumann H., Zillig W.;
RT "Nucleotide sequence of the viral protein TPX of the TTV1 variant
RT VT3.";
RL Nucleic Acids Res. 18:2171-2171(1990).
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; X14717; CAA32838.1; -;

DR PIR; S06686; S06686.
 DR PIR; S15921; S15921.
 KW Repeat.
 FT DOMAIN 278 467 3 THR-PRO REPEATS REGIONS AND TWO NEAR
 FT REPEAT 278 467 IDENTICAL REPEATS.
 FT REPEAT 278 367 THR-PRO (N).
 FT REPEAT 368 377 THR-PRO (N).
 FT REPEAT 378 436 THR-PRO (N).
 FT REPEAT 437 446 THR-PRO (N).
 FT REPEAT 447 467 THR-PRO (N).
 SQ SEQUENCE 474 AA; 49507 MW; 1037109136598829 CRC64;
 Query Match 30.6%; Score 53; DB 1; Length 474;
 Best Local Similarity 42.1%; Pred. No. 9.6;
 Matches 16; Conservative 7; Mismatches 9; Indels 6; Gaps 2;

3 KNTVTAS-----VDPITDLOADGSAALPSAVALTSP 35
 63 QSTITVASGTPNIDPTIALYNNSSY-SNLTLVSP 99

RESULT 8
 YX42_MYCTU STANDARD; PRT; 243 AA.
 AC 053392;
 DT 15-JUL-1999 (Rel. 38, Created)
 DT 15-JUL-1999 (Rel. 38, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Putative methyltransferase RY3342 (EC 2.1.1.1.-).
 GN RY3342 OR MT3445 OR MTV016.42.
 OS Mycobacterium tuberculosis.
 OC Bacteria; Actinobacteria; Actinobacteria (class); Actinobacteridae;
 OC Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium.
 OX NCBI_TaxID=1773;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=H37Rv;
 RA MEDLINE=98295987; PubMed=9634230;
 RA Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,
 RA Gordon S.V., Eiglmeier K., Gas S., Barry C.E. III, Tekala F.,
 RA Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,
 RA Davies R., Devlin K., Felwell T., Gentles S., Hamlin N., Holtroyd S.,
 RA Hornsby T., Jagels K., Krogh A., McLean J., Moule S., Murphy L.,
 RA Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,
 RA Rutter S., Seeger K., Skellern S., Squares S., Squares R.,
 RA Stulson J.E., Taylor K., Whitehead S., Barrrell B.G.;
 RA "Deciphering the biology of Mycobacterium tuberculosis from the
 complete genome sequence.";
 RL Nature 393:537-544(1998).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=CDC 1551 / Oshkosh;
 RA Fleischmann R.D., Alland D., Eisen J.A., Carpenter L., White O.,
 RA Peterson J., Deboy R., Dodson R., Gwinn M.L., Haft D., Hickey E.,
 RA Kolony J.F., Nelson W.C., Umayan L.A., Ermolaeva M.D., Salzberg S.L.,
 RA Delcher A., Utterback T., Weidman J., Khouri H., Gill J., Mikula A.,
 RA Bisht W.;
 RA "Whole genome comparison of Mycobacterium tuberculosis clinical and
 laboratory strains.";
 RL Submitted (APR-2001) to the EMBL/GenBank/DBJ databases.
 CC -1- SIMILARITY: BELONGS TO THE METHYLTRANSFERASE SUPERFAMILY.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation-
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 or send an email to license@isb-sib.ch).

DR Tubercultet; RV3342;
 DR InterPro; IPR000051; SAM_bind.
 KW Hypothetical protein; Transferase; Methyltransferase;
 KW Complete proteome.
 SQ SEQUENCE 243 AA; 26640 MW; BD755GCSAF70C4D3 CRC64;
 Query Match 30.3%; Score 52.5; DB 1; Length 243;
 Best Local Similarity 45.2%; Pred. No. 5.1;
 Matches 14; Conservative 5; Mismatches 11; Indels 1; Gaps 1;

1 VERKNTVTASVDPITDLOADGSAALPSAVAL 31
 57 VERGLDVA-VDPREPMDVLRALPQTVAL 86

RESULT 9
 CSA_DICD1 STANDARD; PRT; 514 AA.
 ID CSA_DICD1
 AC P08796; P19408;
 DT 01-NOV-1988 (Rel. 09, Created)
 DT 01-FEB-1994 (Rel. 28, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Contact site A protein precursor (CSA) (Membrane-associated
 glycoprotein gp80) (Cell adhesion molecule gp80).
 GN CSA.
 OS Dictyostelium discoideum (Slime mold).
 OC Eukaryota; Mycetozoa; Dictyostelidae; Dictyostelium.
 OX NCBI_TaxID=44689;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Noegel A., Gerisch G., Stadler J., Westphal M.;
 RA "Complete sequence and transcript regulation of a cell adhesion
 protein from aggregating Dictyostelium cells.";
 RL EMBO J. 5:1473-1476(1986).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC MEDLINE=89105362; PubMed=3063296;
 RA Siu C.H., Wong L.M., Lam T.Y., Kamboj R.K., Choi A., Cho A.;
 RA "Molecular mechanisms of cell-cell interaction in Dictyostelium
 RT discoideum.";
 RL Biochem. Cell Biol. 66:1089-1099(1988).
 RN [3]
 RP SEQUENCE FROM N.A.
 RC STRAIN=AX2;
 RA MEDLINE=92406924; PubMed=1326559;
 RX Desbates L., Lam T.Y., Wong L.M., Siu C.H.;
 RA "Identification of a unique cAMP-response element in the gene
 RT encoding the cell adhesion molecule gp80 in Dictyostelium
 RT discoideum.";
 RL J. Biol. Chem. 267:19655-19664(1992).
 RN [4]
 RP SEQUENCE OF 20-49.
 RA Wong L.M., Siu C.H.;
 RA "Cloning of cDNA for the contact site A glycoprotein of Dictyostelium
 RT discoideum.";
 RL Proc. Natl. Acad. Sci. U.S.A. 83:4248-4252(1986).
 RN [5]
 RP CELL-BINDING DOMAIN.
 RX MEDLINE=89034443; PubMed=3182938;
 RA Kamboj R.K., Wong L.M., Lam T.Y., Siu C.H.;
 RA "Mapping of a cell-binding domain in the cell adhesion molecule gp80
 RT of Dictyostelium discoideum.";
 RL J. Cell Biol. 107:1835-1843(1988).
 RN [6]
 RP GPI-ANCHOR.
 RX MEDLINE=89251561; PubMed=2721485;
 RA Stadler J., Keenan T.W., Bauer G., Gerisch G.;
 RA "The contact site A glycoprotein of Dictyostelium discoideum carries
 RT a phospholipid anchor of a novel type.";
 RL EMBO J. 8:371-377(1989).
 CC -1- FUNCTION: THIS CELL-SURFACE GLYCOPROTEIN MEDIATES CELL-CELL
 CC BINDING VIA HOMOPHILIC INTERACTION.
 CC -1- SUBCELLULAR LOCATION: ATTACHED TO THE MEMBRANE BY A GPI-ANCHOR

CC THAT CONTAINS A PHOSPHOCERAMIDE MOLETY. SUCH ANCHOR MEDIATES A
 CC FAST AND LONG PERSISTENCE CELL ADHESION OF THE PROTEIN.
 CC DEVELOPMENTAL STAGE: RESTRICTED TO THE AGGREGATION STAGE OF
 CC DEVELOPMENT OF D. DISCOIDEUM.
 CC -1- PTM: ACYLATED, PHOSPHORYLATED ON SERINE AND N-GLYCOSYLATED WITH
 CC TWO TYPES OF OLIGOSACCHARIDE CHAINS.
 CC -1- MISCELLANEOUS: THE EXPRESSION OF THIS STRINGENTLY REGULATED
 CC PROTEIN DURING CELL DEVELOPMENT IS MEDIATED THROUGH CELL-SURFACE
 CC CAMP RECEPTORS.
 CC -1- SIMILARITY: THE C-TERMINAL REGION CONTAINS CLUSTERS OF PROLINE
 CC REGULARLY ALTERNATING WITH A HYDROXYAMINO ACID AND SHOWS
 CC SIMILARITY TO THE HINGE REGION OF IMMUNOGLOBULINS. THIS DOMAIN
 CC MIGHT ACT AS A SPACER TO ELEVATE SITES ACTIVE IN CELL CONTACT
 CC INTO THE EXTRACELLULAR SPACE.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 CC EMBL; X04004; CAA27634.1; -
 CC EMBL; M36545; AAA33212.1; -
 CC EMBL; X66483; CAA47110.1; -
 CC PIR; A23951; A23951.
 CC PIR; A26410; A26410.
 CC PIR; A31643; A31643.
 CC PIR; S22066; S22066.
 CC PIR; A44100; A44100.
 CC Dictyob; D002005; csaA.
 CC InterPro; IPR002909; IPT_TIG.
 CC Pfam; PF01843; TIG; 3.
 CC Cell adhesion; Glycoprotein; Membrane; Phosphorylation; Repeat;
 CC GPI anchor; Lipoprotein; Signal.
 CC
 CC SIGNAL 1 19
 CC CONTACT SITE A PROTEIN.
 CC
 CC FT PROPEP 493 514
 CC REMOVED IN MATURE FORM (POTENTIAL).
 CC FT DOMAIN 20 453
 CC GLOBULAR (POTENTIAL).
 CC FT DOMAIN 454 485
 CC PRO-RICH (HINGE STRUCTURE).
 CC FT DOMAIN 462 479
 CC 2 X 8 AA REPEATS, PRO-RICH.
 CC FT REPEAT 462 469
 CC 1.
 CC FT REPEAT 472 479
 CC 2.
 CC FT CARBOHYD 128 128
 CC N-LINKED (GLCNAC. . .) (POTENTIAL).
 CC FT CARBOHYD 137 137
 CC N-LINKED (GLCNAC. . .) (POTENTIAL).
 CC FT CARBOHYD 207 207
 CC N-LINKED (GLCNAC. . .) (POTENTIAL).
 CC FT CARBOHYD 294 294
 CC N-LINKED (GLCNAC. . .) (POTENTIAL).
 CC FT CARBOHYD 399 399
 CC N-LINKED (GLCNAC. . .) (POTENTIAL).
 CC FT LIPID 492 492
 CC GPI-ANCHOR (POTENTIAL).
 CC FT CONFLICT 216 216
 CC V -> G (IN REF. 1).
 CC
 CC SQ SEQUENCE 514 AA; 53696 MW; 96A1C7CF42FEC096 CRC64;
 CC
 CC Query Match 30.1%; Score 52; DB 1; Length 514;
 CC Best Local Similarity 26.5%; Pred. No. 15;
 CC Matches 9; Conservative 7; Mismatches 18; Indels 0; Gaps 0;
 CC
 CC QY 2 EKNITVTASVDTIDLLQADGSAIPSAVALTYSP 35
 CC
 CC Db 251 EASNTITAKASTGVDMIVLDNQNGQPIFTFYNP 284
 CC
 CC RESULT 10
 CC YHGF_ECOLI
 CC ID YHGF_ECOLI STANDARD; PRT; 773 AA.
 CC AC P46837; P76689;
 CC DT 01-NOV-1995 (Rel. 32, Created)
 CC DT 15-JUL-1999 (Rel. 38, Last sequence update)
 CC DT 16-OCT-2001 (Rel. 40, Last annotation update)
 CC DE Protein YHGF.
 CC GN YHGF OR B1407.
 CC OS Escherichia coli.
 CC OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;

OC Escherichia.
 OX NCBI_TaxID=562;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=K12 / MG1655;
 RX MEDLINE=97426617; PubMed=9278503;
 RA Blattner F.R.; Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,
 RA Riley M., Glasner J.D., Rode C.K., Mayhew G.F.,
 RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,
 RA Mau B., Shao Y.;
 RT "The complete genome sequence of Escherichia coli K-12.";
 RL Science 277:1453-1474(1997).
 RN [2]
 RP IDENTIFICATION BY MASS SPECTROMETRY.
 RX MEDLINE=99420866; PubMed=10491123;
 RA Fountoulakis M., Takacs M.F., Berndt P., Langen H., Takacs B.;
 RT "Enrichment of low abundance proteins of Escherichia coli by
 RT hydroxyapatite chromatography";
 RL Electrophoresis 20:2181-2195(1999).
 CC -1- SIMILARITY: STRONG, TO H1INFLUENZA H10568.
 CC -1- SIMILARITY: CONTAINS 1 SI MOTIF DOMAIN.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 CC EMBL; U18997; AA58204.1; ALT_FRAME.
 CC EMBL; U18997; AA58205.1; ALT_FRAME.
 CC EMBL; AE000416; AAC76432.1; ALT_INIT.
 CC HSSP; P05055; ISRO.
 CC EcoGene; EGI2932; yhgP.
 CC InterPro; IPR003029; S1.
 CC Pfam; PF00575; S1; 1.
 CC SMART; SM00316; S1; 1.
 CC PROSITE; PS0126; S1; 1.
 CC RNA-binding, Complete proteome.
 CC FT DOMAIN 651 720
 CC SI MOTIF.
 CC FT CONFLICT 754 755
 CC QP -> HA (IN REF. 1; AA58205).
 CC SQ SEQUENCE 773 AA; 85119 MW; EA54D9ED952A8229 CRC64;
 CC
 CC Query Match 30.1%; Score 52; DB 1; Length 773;
 CC Best Local Similarity 37.9%; Pred. No. 23;
 CC Matches 11; Conservative 4; Mismatches 14; Indels 0; Gaps 0;
 CC
 CC QY 5 ITVTASVDTIDLLQADGSAIPSAVALTY 33
 CC
 CC Db 117 IATAGLEPLADLLWSDPSHTPEVAAAY 145
 CC
 CC RESULT 11
 CC UVRA_THETH
 CC ID UVRA_THETH STANDARD; PRT; 952 AA.
 CC AC Q56242;
 CC DT 01-NOV-1997 (Rel. 35, Created)
 CC DT 01-NOV-1997 (Rel. 35, Last sequence update)
 CC DT 16-OCT-2001 (Rel. 40, Last annotation update)
 CC DE Exonuclease ABC subunit A.
 CC GN UVRA.
 CC OS Thermus thermophilus.
 CC OC Bacteria; Thermus/Deinococcus group; Deinococci; Thermals;
 CC OC Thermaceae; Thermus.
 CC OX NCBI_TaxID=274;
 CC RN [1]
 CC RP SEQUENCE FROM N.A.
 CC STRAIN=HB8 / ATCC 27634;
 CC RX MEDLINE=96257202; PubMed=8675016;
 CC RA Yamamoto N., Kato R., Kuramitsu S.;
 RT "Cloning, sequencing and expression of the uvra gene from an
 RT extremely thermophilic bacterium, Thermus thermophilus HB8.";

RL Gene 171:103-106(1996).
 CC - FUNCTION: THE ABC EXCISION NUCLEASE IS A DNA REPAIR ENZYME THAT
 CC CATALYZES THE EXCISION REACTION OF UV-DAMAGED NUCLEOTIDE SEGMENTS
 CC PRODUCING OLIGOMERS HAVING THE MODIFIED BASE(S). UVR4 IS AN ATPASE
 CC AND A DNA-BINDING PROTEIN THAT PREFERENTIALLY BINDS SINGLE-
 CC STRANDED OR UV-IRRADIATED DOUBLE-STRANDED DNA (BY SIMILARITY).
 CC SUBUNIT: CONSISTS OF THREE SUBUNITS, UVR4, UVRB AND UVRC.
 CC - SUBCELLULAR LOCATION: Cytoplasmic (By similarity).
 CC - SIMILARITY: BELONGS TO THE ABC TRANSPORTER FAMILY. UVR4 SUBFAMILY.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).

 EMBL: D49911; BAA08652.1; -
 InterPro: IPR003439; ABC_transport.
 DR InterPro: IPR004602; UVR4.
 DR Pfam: PF00005; ABC_tran; 2.
 DR ProDom: PD000006; ABC_transport; 1.
 DR TIGRfam: TIGR00630; Uvr4; 1.
 DR PROSITE: PS00211; ABC_TRANSPORTER; 2.
 KW SOS response; Excision nuclease; DNA repair; ATP-binding; Repeat;
 KM DNA-binding; Zinc-finger.
 FT NP_BIND 31 38 ATP (POTENTIAL).
 FT NP_BIND 643 650 ATP (POTENTIAL).
 FT ZN_FING 253 280 C4-TYPE.
 FT ZN_FING 742 768 C4-TYPE.
 SQ SEQUENCE 952 AA; 105232 MW; FBE2338B45CA96DD CRC64;
 Query Match 30.1%; Score 52; DB 1; Length 952;
 Best Local Similarity 35.6%; Pred. No. 30;
 Matches 16; Conservative 6; Mismatches 11; Indels 12; Gaps 2;
 QY 1 VENKITYTASVPTIDL-----LDGSAALPSAVALTYSP 35
 DB 890 IEHLVDVKTADWYIDLGPEGRGGEIVAEGT--PREVALUTSP 932
 RESULT 12
 YDR6 SCHPO STANDARD; PRT; 438 AA.
 ID YDR6 SCHPO 013741;
 DT 15-JUN-2002 (Rel. 41, Created)
 DT 15-JUN-2002 (Rel. 41, Last sequence update)
 DE Hypothetical RNA-binding protein Cl6E8.06c in chromosome I.
 GN SPAC16E8.06c.
 OS Schizosaccharomyces pombe (Fission Yeast).
 OC Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;
 OC Schizosaccharomycetales; Schizosaccharomycetaceae;
 OC Schizosaccharomycetes.
 NCBI_TaxID=4896;
 GN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=972;
 RX MEDLINE=21848401; PubMed=11859360;
 RA Wood V., Williams R., Rajandream M.A., Lyne M., Lyne R., Stewart A.,
 RA Sgouras J., Peat N., Hayles J., Baker S., Basham D., Bowman S.,
 RA Brooks K., Brown D., Brown S., Chillingworth T., Churcher C.M.,
 RA Collins M., Connor A., Cronin A., Davis P., Felwell T., Fraser A.,
 RA Gentles S., Goble A., Hamlin N., Harris D., Hidalgo J., Hodgson G.,
 RA Holtroyd S., Hornsby T., Howarth S., Huckle E.J., Hunt S., Jagels K.,
 RA James K., Jones L., Mungall K., Murphy L., Niblett D., Odeh C.,
 RA Mooney P., Moule S., Mungall K., Murphy L., Niblett D., Odeh C.,
 RA Oliver K., O'Neill S., Pearson D., Quail M.A., Rabinowitsch E.,
 RA Rutherford K., Rutter S., Saunders D., Seeger K., Sharp S.,
 RA Selton J., Simmonds N., Squares R., Squares S., Stevens K.,
 RA Taylor K., Taylor R.G., Tivey A., Walsh S.V., Warren T., Whitehead S.,
 RA Woodward J., Volckaert G., Aert R., Robben J., Gymnopoulos B.,

RA Weljens I., Vansietele E., Rieger M., Schaefer M., Mueller-Auer S.,
 RA Gabel C., Fuchs M., Filtz C., Holzer E., Moestl D., Hilbert H.,
 RA Bozym K., Langer I., Beck A., Leisch H., Reinhardt R., Pohl T.M.,
 RA Eger P., Zimmermann W., Wedler H., Mambrot R., Punelle B.,
 RA Gaffean A., Cadieu E., Dreano S., Gloux S., Lelaure V., Mortier S.,
 RA Galibert F., Aves S.J., Xiang Z., Hunt C., Moore K., Huist S.M.,
 RA Lucas M., Rochet M., Galliard C., Tallada V.A., Garzon A., Thode G.,
 RA Daga R.R., Cruzado L., Jimenez J., Sanchez M., del Rey F., Benito J.,
 RA Dominguez A., Revuelta J.L., Moreno S., Armstrong J., Forsburg S.L.,
 RA Cerutti L., Lowe T., McCombie M.R., Paulsen J., Potashkin J.,
 RA Shpakovskii G.V., Uesery D., Bartell B.G., Nure P.,
 RT "The genome sequence of Schizosaccharomyces pombe".
 RL Nature 415:871-880(2002).
 CC - SIMILARITY: TO YEAST NOP12.
 CC - SIMILARITY: CONTAINS 2 RNA RECOGNITION MOTIFS (RRM).
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).

 EMBL: Z98529; CAB11047.1; -
 DR InterPro: IPR000504; RNA_rec_mot.
 DR Pfam: PF00076; rrm; 2.
 DR SMART: SMO0360; RRM; 2.
 DR PROSITE: PS00102; RRM; 2.
 DR PROSITE: PS00030; RRM_RNP_1; 1.
 KW Hypothetical protein; RNA-binding; Repeat.
 FT DOMAIN 164 262 RNA-BINDING (RRM) 1.
 FT DOMAIN 270 348 RNA-BINDING (RRM) 2.
 FT DOMAIN 20 23 POLY-SER.
 FT DOMAIN 81 90 POLY-LYS.
 SQ SEQUENCE 438 AA; 49381 MW; 3E943401F95E7C12 CRC64;
 Query Match 28.9%; Score 50; DB 1; Length 438;
 Best Local Similarity 34.6%; Pred. No. 23;
 Matches 9; Conservative 9; Mismatches 8; Indels 0; Gaps 0;
 QY 5 ITVTASVPTIDLDAGSALPSAVALTYSP 30
 DB 19 LSSSNVPTIDLDAGSALPSAVALTYSP 44
 RESULT 13
 PYR5 MOUSE STANDARD; PRT; 481 AA.
 ID PYR5 MOUSE 013439; 099126;
 DT 01-JAN-1990 (Rel. 13, Created)
 DT 15-JUN-2002 (Rel. 41, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE Uridine 5'-monophosphate synthase (UMP synthase) [Includes: Orotate
 DE phosphoribosyltransferase (EC 2.4.2.10) (OPRase); Orotidine 5'-
 DE phosphate decarboxylase (EC 4.1.1.23) (OMPdecase)].
 GN Umps.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 NCBI_TaxID=10090;
 GN [1]
 RP SEQUENCE FROM N.A.
 RA Strusberg R.;
 RL Submitted (Feb-2001) to the EMBL/GenBank/DBJ databases.
 RN [2]
 RP SEQUENCE OF 216-481 FROM N.A.
 RX MEDLINE=86140253; PubMed=2419341;
 RA Ohmsted C.A., Langdon S.D., Chae C.B., Jones M.E.;
 RT "Expression and sequence analysis of a cDNA encoding the orotidine-5'-
 RT monophosphate decarboxylase domain from Ehrlich ascites uridylylate
 RT synthase".
 RT J. Biol. Chem. 261:4276-4282(1986).

```

CC -!- CATALYTIC ACTIVITY: Orotidine 5'-phosphate + diphosphate = orotate
CC + 5-phospho-alpha-D-ribose 1-diphosphate.
CC -!- PATHWAY: Pyrimidine biosynthesis; fifth and sixth (last) steps.
CC -!- SIMILARITY: IN THE N-TERMINAL SECTION; BELONGS TO THE
CC PURINE/PYRIMIDINE PHOSPHORIBOSYLTRANSFERASE FAMILY.
CC -!- SIMILARITY: IN THE C-TERMINAL SECTION; BELONGS TO THE
CC GMP DECARBOXYLASE FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announcement/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: BC003887; AA03887.1;
DR EMBL: M29195; AAA39859.1;
DR PIR: A25323; DCM5QP.
DR HSP: P03962; DQW.
DR MGD: MG1:1298188; Umps.
DR InterPro: IPR001754; OMPDecase.
DR InterPro: IPR002375; Pp/pv rp transf.
DR InterPro: IPR003816; PRTtransferase.
DR InterPro: IPR004467; Pyre.
DR Pfam: PF00215; OMPDecase; 1.
DR TIGRfam: TIGR00336; pyre; 1.
DR PROSITE: PS00156; OMPDecase; 1.
DR PROSITE: PS00103; PUR_PYR_TRANSFER; 1.
KW Pyrimidine biosynthesis; Multifunctional enzyme; Transferase;
KW Glycosyltransferase; Lyase; decarboxylase.
FT DOMAIN 1 214 CPTASE.
FT DOMAIN 215 220 DOMAIN LINKER.
FT DOMAIN 221 481 OMPDecase.
FT ACT_SITE 314 314 BY SIMILARITY.
SQ SEQUENCE 481 AA; 52292 MW; 761C7EF7960C9C59 CRC64;

Query Match 28.9%; Score 50; DB 1; Length 481;
Best Local Similarity 35.5%; Pred. No. 25;
Matches 11; Conservative 8; Mismatches 12; Indels 0; Gaps 0;

QY 1 VKNITVTASVPTIDLLQAGSALPSAVAL 31
DQ 122 IEDVVTGASVLEIVLQKGLKVTDAIVL 152

RESULT 14
YAM6 SCHPO STANDARD; PRT; 849 AA.
AC Q10061;
DT 01-FEB-1996 (Rel. 33, Created)
DT 01-FEB-1996 (Rel. 33, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Heat shock protein 70 homolog precursor.
GN SPAT1P5.06.
OS Schizosaccharomyces pombe (Fission yeast).
OC Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;
OC Schizosaccharomycetales; Schizosaccharomycetaceae;
OC Schizosaccharomycetes.
OX NCBI_TaxID=4996;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN 972;
EX MEDLINE:11848401; PubMed:11859360;
RA Wood V., Gwilliam R., Rajandream M.A., Lyne M., Lyne R., Stewart A.,
RA Sgouros J., Peat N., Hayles J., Baker S., Basham D., Bowman S.,
RA Brooks K., Brown D., Brown S., Chillingworth T., Churcher C.M.,
RA Collins M., Connor R., Cronin A., Davis P., Feltwell T., Fraser A.,
RA Gentles S., Goble A., Hamlin N., Harris D., Hildalgo J., Hodgson G.,
RA Holtova S., Hornsby T., Howarth S., Huckle E.J., Hunt S., Jagels K.,
RA James K., Jones L., Jones M., Leather S., McDonald S., McLean J.,

```

```

RA Mooney P., Moule S., Mungall K., Murphy L., Niblett D., Odell C.,
RA Oliver K., O'Neil S., Pearson D., Quail M.A., Rabinowitsch E.,
RA Rutherford K., Rutter S., Saunders D., Seeger K., Sharp S.,
RA Skelton J., Simmonds M., Squares R., Squares S., Stevens K.,
RA Taylor K., Taylor R.G., Tivey A., Walsh S.V., Warren T., Whitehead S.,
RA Woodward J., Volckaert G., Aert R., Robben J., Grymopiez B.,
RA Weltjens I., Vanstreels E., Rieger M., Schaefer M., Mueller Auer S.,
RA Gabel C., Fuchs M., Fritz C., Holzer E., Moestl D., Hilbert H.,
RA Borzym K., Langer I., Beck A., Leinhardt H., Wambutt R., Purnelle B.,
RA Eger P., Zimmermann W., Wedler H., Wambutt R., Purnelle B.,
RA Goffeau A., Cadieu E., Dreano S., Gloux S., Lelaue V., Mottier S.,
RA Gallibert F., Aves S.J., Xiang Z., Hunt C., Moore K., Hurst S.M.,
RA Lucas M., Rochet M., Gaillardin C., Tallada V.A., Garzon A., Thode G.,
RA Daga R.R., Cruzado J., Jimenez J., Sanchez M., del Rey P., Benito J.,
RA Dominguez A., Revuelta J.L., Moreno S., Armstrong J., Potsburg S.L.,
RA Cerutti L., Lowe T., McCombie W.R., Paulsen I., Potashkin J.,
RA Shpakovski G.V., Ussery D., Barrell B.G., Nurse P.;
RT "The genome sequence of Schizosaccharomyces pombe.";
RL Nature 415:871-880(2002).
CC -!- SUBCELLULAR LOCATION: Endoplasmic reticulum lumen (Potential).
CC -!- SIMILARITY: BELONGS TO THE HEAT SHOCK PROTEIN 70 FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announcement/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: Z68136; CAA92234.1;
DR InterPro: IPR000886; ER-target.
DR InterPro: IPR001023; HSP70; 1.
DR Pfam: PF00012; HSP70; 1.
DR ProDom: PD000089; HSP70; 1.
DR PROSITE: PS00014; ER_TARGET; 1.
DR PROSITE: PS00297; HSP70_1; FALSE_NEG.
DR PROSITE: PS00329; HSP70_2; 1.
DR PROSITE: PS01036; HSP70_3; 1.
KW Hypothetical protein; Endoplasmic reticulum; Signal; ATP-binding.
FT SIGNAL 1 21 POTENTIAL.
FT CHAIN 22 848 HEAT SHOCK PROTEIN 70 HOMOLOG.
FT SITE 845 848 PREVENT SECRETION FROM ER (POTENTIAL).
SQ SEQUENCE 848 AA; 94897 MW; A1963FD4253F36F3 CRC64;

Query Match 28.9%; Score 50; DB 1; Length 848;
Best Local Similarity 27.3%; Pred. No. 50;
Matches 9; Conservative 11; Mismatches 13; Indels 0; Gaps 0;

QY 1 VERNITVTASVPTIDLLQAGSALPSAVALTY 33
DQ 202 IEVSLRSFSTDPYNIIVDSGSGSTSATVISF 214

RESULT 15
W7DT_HUMAN STANDARD; PRT; 925 AA.
ID W70T_HUMAN
AC P57737;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE 70 kDa WD-repeat tumor rejection antigen homolog.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA Kawabata A., Hiki J., Kobatake N., Inagaki H., Ikema Y., Okamoto S.,
RA Okitani R., Oka T., Suzuki Y., Obayashi M., Nishi T., Shibahara T.,
RA Tanaka T., Nakamura Y., Isogai T., Sugano S.;
RT "NEDO human cDNA sequencing project.";
RL Submitted (AUG-2000) to the EMBL/GenBank/DBJ databases.

```



```

DE precursor;
GN YJL079C OR J1022.
OS Saccharomyces cerevisiae (Baker's Yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
CX NCHI_TaxID:4932;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=5788C;
RX MEDLINE=96093911; PubMed=7483841;
RA Miosga T., Schaaf-Gerstenschlaeger I., Chalarz N., Baur A.,
RA Boles E., Fournier C., Schmitt S., Velten C., Wilhelm N.,
RA Zimmermann F.K.;
RT "Sequence analysis of a 33.1 kb fragment from the left arm of
RT Saccharomyces cerevisiae chromosome X, including putative proteins
RT with leucine zippers, a fungal Zn(II)-Cys6 binuclear cluster domain
RT and a putative alpha 2-SCB-alpha 2 binding site.";
RL Yeast 11:681-689(1995).
CC - SIMILARITY: BELONGS TO A FAMILY THAT GROUPS MAMMALIAN SCP/TPX1;
CC - INSECTS AG/AGS; FUNGI SC7/SC14 AND PLANTS PR-1.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; Z49154; CAA89372.1; -
DR EMBL; X83502; CAA5491.1; -
DR EMBL; X88851; CAA61315.1; -
DR HSSP; PC4284; 1CPE.
DR SGD; S0003615; YJL079C.
DR InterPro; IPR001283; Allrgn_V5/Tpx1.
DR Pfam; PF00188; SCP_1.
DR PRINTS; PR00837; V5TPXLIKE.
DR ProDom; PD000542; Allrgn_V5/Tpx1; 1.
DR SMART; SM00198; SCP_1.
DR PROSITE; PS01009; SCP_AG5_PRI_SC1_1; 1.
DR PROSITE; PS01010; SCP_AG5_PRI_SC2_2; 1.
KW Hypothetical protein; Signal; POTENTIAL.
FT SIGNAL 1 19
FT CHAIN 20 299 HYPOTHETICAL PROTEIN YJL079C.
FT DOMAIN 102 165 ALA/SER/THR-RICH.
FT SEQUENCE 299 AA; 30634 MW; A316769CA87C5679 CRC64;
Query Match 28.3%; Score 49; DB 1; Length 299;
Best Local Similarity 31.4%; Pred. No. 20;
Matches 11; Conservative 9; Mismatches 15; Indels 0; Gaps 0;
QY 2 EKNITVTSVDPPTDLDQAGSALPSAVALTSPA 36
DB 104 DSTTLLSSSTSSQSLAQTTSIPAAASTTSTA 138
RESULT 18
UDG_STRN
ID UDG_STRN STANDARD; PRT; 394 AA.
AC Q57346; Q54610; P72519;
DT 01-NOV-1997 (Rel. 35, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE UDP-glucose 6 dehydrogenase (EC 1.1.1.22) (UDP-Glc dehydrogenase)
DE (UDP-GlcDH) (UDPGDH).
GN CAP1A OR CFS3D.
OS Streptococcus pneumoniae.
OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
OC Streptococcus.
CX NCHI_TaxID:1113;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=406 / Type 3;

```

```

RX MEDLINE=95014083; PubMed=7929009;
RA Arrecubieta C., Lopez R., Garcia E.;
RT "Molecular characterization of cap3A, a gene from the operon required
RT for the synthesis of the capsule of Streptococcus pneumoniae type 1;
RT sequencing of mutations responsible for the unencapsulated phenotype
RT and localization of the capsular cluster on the pneumococcal
RT chromosome.";
RL J. Bacteriol. 176:6375-6383(1994).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=406 / Type 3;
RX MEDLINE=96144241; PubMed=8566758;
RA Arrecubieta C., Garcia E., Lopez R.;
RT "Sequence and transcriptional analysis of a DNA region involved in
RT the production of capsular polysaccharide in Streptococcus pneumoniae
RT type 3.";
RL Gene 167:1-7(1995).
RN [3]
RP REVISIONS TO 129 AND 228.
RA Garcia E., Arrecubieta C., Munoz R., Mollerach M., Lopez R.;
RL Submitted (SEP-1996) to the EMBL/GenBank/DBSJ databases.
RN [4]
RP SEQUENCE FROM N.A.
RC STRAIN=WU2 / Serotype 3;
RX MEDLINE=95173611; PubMed=7869055;
RA Dillard J.P., Vandersea M.W., Yother J.;
RT "Characterization of the cassette containing genes for type 3
RT capsular polysaccharide biosynthesis in Streptococcus pneumoniae.";
RL J. Exp. Med. 181:973-983(1995).
CC - FUNCTION: CATALYZES THE FORMATION OF UDP-GLUCURONIC ACID WHICH IS
CC REQUIRED FOR CAPSULAR HYALURONIC ACID SYNTHESIS. DIRECTLY
CC RESPONSIBLE FOR THE TRANSFORMATION OF SOME UNENCAPSULATED
CC SEROTYPE-3 SP MUTANTS TO THE ENCAPSULATED PHENOTYPE.
CC - CATALYTIC ACTIVITY: UDP-glucose + 2 NAD(+) + H(2)O -> UDP
CC glucuronate + 2 NADH.
CC - PATHWAY: HYALURONATE SYNTHESIS.
CC - SIMILARITY: BELONGS TO THE UDP-GLUCOSE/GDP-MANNOSE DEHYDROGENASES
CC FAMILY.

```

This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (see <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

```

DR EMBL; Z47210; CAA87403.1; -
DR EMBL; Z12159; CAA78147.1; -
DR EMBL; U15171; AAC43311.1; -
DR InterPro; IPR001732; UDEG_MGDP dh.
DR Pfam; PF00984; UDEG_MGDP dh; 1.
DR Pfam; PF03720; UDEG_MGDP dh C; 1.
DR Pfam; PF03721; UDEG_MGDP dh_N; 1.
KW Oxidoreductase; NAD.
FT NP_BIND 2 19
FT ACT_SITE 259 259 BY SIMILARITY.
FT CONFLICT 215 215 R -> A (IN REF. 4).
FT CONFLICT 247 247 Y -> D (IN REF. 4).
SQ SEQUENCE 394 AA; 44697 MW; C68671ED88140FC3 CRC64;

```

Query Match 28.3%; Score 49; DB 1; Length 394;
Best Local Similarity 27.7%; Pred. No. 28;
Matches 13; Conservative 8; Mismatches 8; Indels 18; Gaps 1;

```

QY 1 VEKNITVTSVDPPTDLDQAGSALPSAV 29
DB 55 VEKELNLEASLDFAHYKQVEXAIATPTNYDVLNQDFTSSVEAAI 101

```

```

RESULT 19
EXU_DROVI
ID EXU_DROVI STANDARD; PRT; 486 AA.

```

AC Q24747;
 DT 01-NOV-1997 (Rel. 35, Created)
 DT 01-NOV-1997 (Rel. 35, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Maternal exuperantia protein.
 GN EXU.
 OS Drosophila virilis (Fruit Fly).
 OC Eukaryota; Metazoa; Arthropoda; Mandibulata; Pancrustacea; Hexapoda;
 CC Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
 CC Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
 NC NCBI_TaxID:7244;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA MEDLINE=94350208; PubMed=8070663;
 RT Luk S.K.-S., Kilpatrick M., Kerr K., Macdonald P.M.;
 RT "Components acting in localization of bicoid mRNA are conserved among
 RT Drosophila species";
 RT Genetics 137:521-530(1994).
 CC -1- FUNCTION: ENSURES THE PROPER LOCALIZATION OF THE MRNA OF THE
 CC BICOID GENE TO THE ANTERIOR REGIONS OF THE OOCYTE THUS PLAYING
 CC A FUNDAMENTAL ROLE IN THE ESTABLISHMENT OF THE POLARITY OF THE
 CC OOCYTE. MAY BIND THE BCD MRNA (BY SIMILARITY).
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 CC EMBL: L22555; AAA28525.1; -
 CC DR FLYBASE: FBgn0013112; Dvnl\exu.
 CC KM Developmental protein; RNA-binding.
 CC SQ SEQUENCE 486 AA; 53957 MW; FD30313AAED0309A CRC64;
 CC
 CC Query Match 28.3%; Score 49; DB 1; Length 486;
 CC Best Local Similarity 44.0%; Pred. No. 35;
 CC Matches 11; Conservative 3; Mismatches 11; Indels 0; Gaps 0;
 CC
 CC 7 VTASVDPITDILQMGSLASVAL 31
 CC 2 VAVSIDVIDISADOSTAASGVAI 26
 CC
 CC RESULT 20
 CC PK7_HUMAN STANDARD; PRT; 1070 AA.
 AC Q13308; Q13417;
 DT 30-MAY-2000 (Rel. 39, Created)
 DT 30-MAY-2000 (Rel. 39, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE Tyrosine-protein kinase-like 7 precursor (Colon carcinoma kinase-4)
 DE (CKK-4).
 OS Homo sapiens (Human).
 GN PK7 OR CKK4.
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 CC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 NC NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Colon carcinoma, and Placenta;
 RX MEDLINE=96074849; PubMed=7478540;
 RA Mossie K., Jallat B., Alves P., Sures I., Ploewman G.D., Ullrich A.;
 RT "Colon carcinoma kinase-4 defines a new subclass of the receptor
 RT tyrosine kinase family";
 RT Oncogene 11:2179-2184(1995).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Fibroblast;
 RX MEDLINE=97037064; PubMed=8882711;
 RA Park S.-K., Lee H.-S., Lee S.-T.;
 RT "Characterization of the human full-length PK7 cDNA encoding a

RT receptor protein tyrosine kinase-like molecule closely related to
 RT chick Klg.";
 RL J. Biochem. 119:235-239(1996).
 RN [3]
 RP SEQUENCE FROM N.A., AND REVISION TO 834.
 RA Lee S.-T., Park S.-K., Lee H.-S., Ji A.-R., Jung J.W.;
 RL Submitted (Nov-2001) to the EMBL/GenBank/DBJ databases.
 CC -1- FUNCTION: MAY FUNCTION AS A CELL ADHESION MOLECULE. LACKS PROBABLY
 CC THE CATALYTIC ACTIVITY OF TYROSINE KINASE. MAY BE CONNECTED TO THE
 CC PATHOPHYSIOLOGY OF COLON CARCINOMAS AND/OR MAY REPRESENT A TUMOR
 CC PROGRESSION MARKER.
 CC
 CC -1- SUBCELLULAR LOCATION: Type I membrane protein.
 CC
 CC -1- TISSUE SPECIFICITY: HIGHLY EXPRESSED IN LUNG, LIVER, PANCREAS,
 CC KIDNEY, PLACENTA AND MELANOCYTES. WEAKLY EXPRESSED IN THYROID
 CC GLAND, OVARY, BRAIN, HEART AND SKELETAL MUSCLE. ALSO EXPRESSED IN
 CC ERYTHROLEUKEMIA CELLS. BUT NOT EXPRESSED IN COLON.
 CC
 CC -1- SIMILARITY: BELONGS TO THE TYR FAMILY OF PROTEIN KINASES. INSULIN
 CC RECEPTOR SUBFAMILY.
 CC
 CC -1- SIMILARITY: CONTAINS 7 IMMUNOGLOBULIN-LIKE C2-TYPE DOMAINS.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 CC EMBL: U33635; AAA87565.1; -
 CC DR EMBL: U40271; AAC50484.2; -
 CC DR EMBL: AF447176; AAL39062.1; JOINED.
 CC DR EMBL: AF447157; AAL39062.1; JOINED.
 CC DR EMBL: AF447158; AAL39062.1; JOINED.
 CC DR EMBL: AF447162; AAL39062.1; JOINED.
 CC DR EMBL: AF447164; AAL39062.1; JOINED.
 CC DR EMBL: AF447167; AAL39062.1; JOINED.
 CC DR EMBL: AF447170; AAL39062.1; JOINED.
 CC DR EMBL: AF447171; AAL39062.1; JOINED.
 CC DR EMBL: AF447173; AAL39062.1; JOINED.
 CC DR EMBL: AF447174; AAL39062.1; JOINED.
 CC DR EMBL: AF447175; AAL39062.1; JOINED.
 CC DR HSP; P08631; IAD5.
 CC DR Genew; HGNC:9618; PK7.
 CC
 CC MIM: 601890; -
 CC DR InterPro; IPR000719; Euk_Pkinase.
 CC DR InterPro; IPR003006; Ig_MHC.
 CC DR InterPro; IPR003598; Ig_C2.
 CC DR InterPro; IPR003600; Ig_Like.
 CC DR InterPro; IPR020211; RTKinaseI.
 CC DR InterPro; IPR001245; Tyr_pkinase.
 CC DR Pfam; PF00047; Ig_7.
 CC DR Pfam; PF00069; Pkinase; 1.
 CC DR PRINTS; PR00109; TYRKINASE.
 CC DR ProDom; PD000001; Euk_Pkinase; 1.
 CC DR SMART; SM00410; IG_Like; 2.
 CC DR SMART; SM00408; IGC2; 5.
 CC DR SMART; SM00219; TYRK; 1.
 CC DR PROSITE; PS00109; PROTEIN_KINASE_TYR; 1.
 CC DR PROSITE; PS00011; PROTEIN_KINASE_DOM; 1.
 CC DR PROSITE; PS00239; RECEPTOR_TYR_KIN_II; FALSE NEG.
 CC KW Receptor; Transmembrane; Signal; Glycoprotein; Cell adhesion;
 CC Immunoglobulin domain; Repeat.
 CC
 CC SIGNAL 1 30
 CC
 CC FT CHAIN 31 1070 TYROSINE-PROTEIN KINASE-LIKE 7.
 CC FT DOMAIN 31 704 EXTRACELLULAR (POTENTIAL).
 CC FT TRANSMEM 705 725 POTENTIAL.
 CC FT DOMAIN 726 1070 CYTOPLASMIC (POTENTIAL).
 CC FT DOMAIN 46 108 IG-LIKE C2-TYPE DOMAIN 1.
 CC FT DOMAIN 143 208 IG-LIKE C2-TYPE DOMAIN 2.
 CC FT DOMAIN 239 308 IG-LIKE C2-TYPE DOMAIN 3.
 CC FT DOMAIN 336 398 IG-LIKE C2-TYPE DOMAIN 4.
 CC FT DOMAIN 426 488 IG-LIKE C2-TYPE DOMAIN 5.
 CC FT DOMAIN 517 577 IG-LIKE C2-TYPE DOMAIN 6.

```

PT DOMAIN 606 671 IG-LIKE C2-TYPE DOMAIN 7.
PT DOMAIN 796 1066 PROTEIN KINASE; INACTIVE.
PT DISULFID 53 101 BY SIMILARITY.
PT DISULFID 150 200 BY SIMILARITY.
PT DISULFID 246 301 BY SIMILARITY.
PT DISULFID 343 391 BY SIMILARITY.
PT DISULFID 433 481 BY SIMILARITY.
PT DISULFID 524 570 BY SIMILARITY.
PT DISULFID 613 664 BY SIMILARITY.
PT CARBOHYD 116 116 N-LINKED (GLCNAC. .) (POTENTIAL).
PT CARBOHYD 175 175 N-LINKED (GLCNAC. .) (POTENTIAL).
PT CARBOHYD 184 184 N-LINKED (GLCNAC. .) (POTENTIAL).
PT CARBOHYD 214 214 N-LINKED (GLCNAC. .) (POTENTIAL).
PT CARBOHYD 268 268 N-LINKED (GLCNAC. .) (POTENTIAL).
PT CARBOHYD 283 283 N-LINKED (GLCNAC. .) (POTENTIAL).
PT CARBOHYD 405 405 N-LINKED (GLCNAC. .) (POTENTIAL).
PT CARBOHYD 463 463 N-LINKED (GLCNAC. .) (POTENTIAL).
PT CARBOHYD 567 567 N-LINKED (GLCNAC. .) (POTENTIAL).
PT CARBOHYD 646 646 N-LINKED (GLCNAC. .) (POTENTIAL).
PT CONFLICT 92 92 P -> R (IN REF. 2 AND 3).
PT CONFLICT 147 147 K -> T (IN REF. 2 AND 3).
PT CONFLICT 207 207 S -> G (IN REF. 2 AND 3).
PT CONFLICT 495 496 VL -> RV (IN REF. 2 AND 3).
PT CONFLICT 515 515 G -> E (IN REF. 2 AND 3).
PT CONFLICT 881 881 E -> G (IN REF. 2 AND 3).
PT CONFLICT 969 969 A -> P (IN REF. 2 AND 3).
PT CONFLICT 992 992 S -> F (IN REF. 2 AND 3).
SQ SEQUENCE 1070 AA; 47CDF25B8E3698A5 CRC64;

Query Match 28.3%; Score 49; DB 1; Length 1070;
Best Local Similarity 43.8%; Pred. No. 90;
Matches 14; Conservative 3; Mismatches 11; Indels 4; Gaps 1;

QY 2 EKNTVTASV ---DPTDLDLQAGSLPSAV 29
DB 517 DKHATVPCATGRKXETIKWKAQSSLPWV 548

RESULT 21
SPCO MOUSE
ID SPCO MOUSE STANDARD; PRT; 2363 AA.
AC Q62261;
DT 01 NOV-1997 (Rel. 35, Created)
DT 01 NOV-1997 (Rel. 35, Last annotation update)
DT 15 JUN-2002 (Rel. 41, Last annotation update)
DE Spectrin beta chain, brain 1 (Spectrin, non-erythroid beta chain 1)
DE (beta II spectrin) (Fodrin beta chain).
UN SPTBN1 OR SPTB2 OR SPNB2 OR SPNB-2.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.
OX NCBI_TaxID=10090;
RN [1] TuxID=10090;
RP SEQUENCE FROM N.A.
RC STRAIN-BALB/C;
RX MEDLINE=93240985; PubMed=8479293;
RA Ma Y., Zimmer W.E., Riederer B.M., Goodman S.R.;
RT "The complete amino acid sequence for brain beta spectrin (beta
fodrin): relationship to globin sequences."
RL Brain Res. Mol. Brain Res. 18:87-99(1993).
[2]
RP STRUCTURE BY NMR OF 2199-2304.
RX MEDLINE=94269558; PubMed=8208297;
RA Macias M.J., Musacchio A., Ponstingl H., Nilges M., Saraste M.,
RA Oschkinat H.;
RT "Structure of the spectrin homology domain from beta-spectrin."
RL Nature 369:675-677(1994).
[3]
RP STRUCTURE BY NMR OF 2199-2304.
RX MEDLINE=9734712; PubMed=9199409;
RA Nilges M., Macias M.J., O'Donoghue S.I., Oschkinat H.;
RT "Automated NMR interpretation with ambiguous distance restraints:
the refined NMR solution structure of the spectrin homology domain
```

```

RT from beta-spectrin.";
RL J. Mol. Biol. 269:408-422(1997).
[4]
RN X-RAY CRYSTALLOGRAPHY (2.0 ANGSTROMS) OF 2199-2304.
RX MEDLINE=96030773; PubMed=7588597;
RA Hyvoenen M., Macias M.J., Nilges M., Oschkinat H., Saraste M.,
RA Wilmanns M.;
RT "Structure of the binding site for inositol phosphates in a PH
domain.";
RL EMBO J. 14:4676-4681(1995).
CC -!- FUNCTION: FODRIN, WHICH SEEMS TO BE INVOLVED IN SECRETION,
INTERACTS WITH CALMODULIN IN A CALCIUM-DEPENDENT MANNER AND IS
THUS CANDIDATE FOR THE CALCIUM-DEPENDENT MOVEMENT OF THE
CYTOSKELETON AT THE MEMBRANE.
CC -!- SUBUNIT: LIKE ERYTHROCYTE SPECTRIN, THE SPECTRIN-LIKE PROTEINS
ARE CAPABLE TO FORM DIMERS WHICH CAN FURTHER ASSOCIATE TO
TETRAMERS.
CC -!- SIMILARITY: BELONGS TO THE SPECTRIN FAMILY.
CC -!- SIMILARITY: CONTAINS 2 CALPONIN-HOMOLOGY (CH) DOMAINS.
CC -!- SIMILARITY: CONTAINS 1 PH DOMAIN.
CC -!- SIMILARITY: CONTAINS 17 SPECTRIN REPEATS.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation
at the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@sib-sib.ch).
DR EMBL; M74773; AAC42040.1;
DR PDB; 1BTN; 08-MAR-96.
DR PDB; 1MPH; 16-JUN-97.
DR MGD; MGI:98388; Spnb2.
DR InterPro; IPR001589; Actbind actinin.
DR InterPro; IPR001715; Calponin-like.
DR InterPro; IPR001849; PH.
DR InterPro; IPR02017; Spectrin_PH.
DR InterPro; IPR001605; Spectrin_PH.
DR Pfam; PF00169; PH; 1.
DR Pfam; PF00307; CH; 2.
DR Pfam; PF00435; spectrin; 18.
DR PRINTS; PR00683; SPECTRINPH.
DR SMART; SM00033; CH; 2.
DR SMART; SM00233; PH; 1.
DR SMART; SM00150; SPEC; 16.
DR PROSITE; PS00019; ACTININ 1; 1.
DR PROSITE; PS00020; ACTININ 2; 1.
DR PROSITE; PS00021; CH; 2.
DR PROSITE; PS00003; PH_DOMAIN; 1.
KW Cytoskeleton; Membrane; Repeat; Actin-binding; Capping protein;
KW 3D-structure.
FT DOMAIN 1 275 ACTIN-BINDING (BY SIMILARITY).
FT DOMAIN 54 158 CH 1.
FT DOMAIN 173 275 CH 2.
FT REPEAT 276 384 SPECTRIN 1.
FT REPEAT 385 498 SPECTRIN 2.
FT REPEAT 499 608 SPECTRIN 3.
FT REPEAT 609 714 SPECTRIN 4.
FT REPEAT 715 819 SPECTRIN 5.
FT REPEAT 820 925 SPECTRIN 6.
FT REPEAT 926 1032 SPECTRIN 7.
FT REPEAT 1033 1139 SPECTRIN 8.
FT REPEAT 1140 1245 SPECTRIN 9.
FT REPEAT 1246 1350 SPECTRIN 10.
FT REPEAT 1351 1462 SPECTRIN 11.
FT REPEAT 1463 1562 SPECTRIN 12.
FT REPEAT 1563 1668 SPECTRIN 13.
FT REPEAT 1669 1775 SPECTRIN 14.
FT REPEAT 1776 1881 SPECTRIN 15.
FT REPEAT 1882 1987 SPECTRIN 16.
FT REPEAT 1988 2132 SPECTRIN 17.
FT DOMAIN 2196 2306 PH.
```

SQ SEQUENCE 2363 AA; 274420 MW; 64C9E4BD26BECT78 CRC64;
 Query Match 28.0%; Score 48.5; DB 1; Length 2363;
 Best Local Similarity 42.4%; Pred. No. 2.7e+02;
 Matches 14; Conservative 3; Mismatches 13; Indels 3; Gaps 1;

QY 1 VERNITVTASVD---PTIDLDGSAALPSAVA 30
 DB 783 VKKHQDAEITVYRPIDTLEHQASALPQAH 815

RESULT 22
 YH74_MERTH STANDARD; PRT; 172 AA.
 ID YH74_MERTH STANDARD; PRT; 172 AA.
 AC 027802;
 DT 15-JUL-1998 (Rel. 36, Created)
 DT 15-JUL-1998 (Rel. 36, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 Hypothetical protein M01774.
 MTH1774.
 OS Methanobacterium thermoautotrophicum.
 OC Archaea; Euryarchaeota; Methanobacteriales;
 OC Methanobacteriaceae; Methanothermobacter.
 OX NCBI_TaxID=187420;
 RN (1)
 RP SEQUENCE FROM N.A.
 RC STRAIN=Delta H;
 RX MEDLINE=98037514; PubMed=9371463;
 RA Smith D.R., Doucette-Stamm L.A., Doughtery C., Lee H.-M., Dubois J.,
 RA Aldredge T., Bashirzadeh R., Blakely D., Cook R., Gilbert K.,
 RA Harrison D., Hoang L., Keagle P., Lum W., Pochler B., Qiu D.,
 RA Spadafora R., Vicare R., Wang Y., Wierzbowski J., Gibson R.,
 RA Jiwani N., Caruso A., Bush D., Safer H., Patwell D., Prabhakar S.,
 RA McDougall S., Shimer G., Goyal A., Pietrowski S., Church G.M.,
 RA Daniels C.J., Mao J.-I., Rice P., Noelling J., Reeve J.N.;
 RT "Complete genome sequence of Methanobacterium thermoautotrophicum
 RT deltaH: functional analysis and comparative genomics."
 RL J. Bacteriol. 179:7135-7155(1997).
 CC -1- SIMILARITY: BELONGS TO THE UPF0025 FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).

 DR EMBL: AE000932; AAB86240.1; -;
 DR InterPro: IPR004843; M-peptidase.
 DR InterPro: IPR004844; S/T phosphatase.
 DR InterPro: IPR000979; UPF0025.
 DR Pfam: PF00149; Metallophos; 1.
 DR TIGRFAMs: TIGR00040; UPF0025; 1.
 DR PROSITE: PS01269; UPF0025; 1.
 KW Hypothetical protein; Complete proteome.
 SQ SEQUENCE 172 AA; 19223 MW; F59AB11A5218BF2 CRC64;

Query Match 27.7%; Score 48; DB 1; Length 172;
 Best Local Similarity 33.3%; Pred. No. 14;
 Matches 10; Conservative 7; Mismatches 13; Indels 0; Gaps 0;

QY 6 TWTASVPTIDLDGSAALPSAVALTSP 35
 DB 134 TTPRLTDSVWVLRIDSEKLDALRTGAP 163

RESULT 23
 MRAM_MYCPU STANDARD; PRT; 239 AA.
 ID MRAM_MYCPU STANDARD; PRT; 239 AA.
 AC 098075;
 DT 15-JUN-2002 (Rel. 41, Created)
 DT 15-JUN-2002 (Rel. 41, Last sequence update)

DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE S-adenosyl-methyltransferase mrav (EC 2.1.1.-).
 RN MRAM OR MYPU 4930.
 OS Mycoplasma pulmonis.
 OC Bacteria; Firmicutes; Mollicutes; Mycoplasmataceae; Mycoplasma.
 OX NCBI_TaxID=2107;
 RN (1)
 RP SEQUENCE FROM N.A.
 RC STRAIN=UAB CTIP;
 RX MEDLINE=21257165; PubMed=11353084;
 RA Chamberaud I., Helling R., Ferris S., Barde V., Samson D., Galisson F.,
 RA Moszer I., Dybvig K., Wroblewski H., Viart A., Rocha E.P.C.,
 RA Blanchard A.;
 RT "The complete genome sequence of the murine respiratory pathogen
 RT Mycoplasma pulmonis."
 RL Nucleic Acids Res. 29:2145-2153(2001).
 CC -1- FUNCTION: Exhibits a S-adenosyl dependent methyltransferase
 CC activity (by similarity).
 CC -1- SIMILARITY: BELONGS TO THE MRAM FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).

 DR EMBL: AL445564; CAC13666.1; -;
 DR Mypulist; MYPU_4930; -;
 DR InterPro: IPR002903; Bac_MetTrfase.
 DR Pfam: PF01795; Methyltransf 5; 1.
 DR ProDom: PD04685; Bac MetTrfase; 1.
 DR TIGRFAMs: TIGR00006; UPF0117; 1.
 KW Transferase; Methyltransferase; Complete proteome.
 SQ SEQUENCE 299 AA; 34245 MW; 2E0D65FD3FF6DCP CRC64;

Query Match 27.7%; Score 48; DB 1; Length 299;
 Best Local Similarity 30.6%; Pred. No. 27;
 Matches 11; Conservative 9; Mismatches 16; Indels 0; Gaps 0;

QY 1 VERNITVTASVPTIDLDGSAALPSAVALTSP 36
 DB 163 VAKGIVNRPDKTLELVIRSSLPALVRKNDA 198

RESULT 24
 MTC2_CHVP1 STANDARD; PRT; 326 AA.
 ID MTC2_CHVP1 STANDARD; PRT; 326 AA.
 AC P31119; Q84569;
 DT 01-JUL-1993 (Rel. 26, Created)
 DT 15-DEC-1998 (Rel. 37, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Modification methylase CviA11 (EC 2.1.1.72) (adenine-specific
 DE methyltransferase CviA11) (M_CviA11).
 GN CviA11 OR A251R.
 OS Paramesidium bursaria Chlorella virus 1 (PBCV-1).
 OC Viruses; dsDNA viruses, no RNA stage; Phycodnaviridae; Chlorevirus.
 OX NCBI_TaxID=10506;
 RN (1)
 RP SEQUENCE FROM N.A.
 RX MEDLINE=93065201; PubMed=1437552;
 RA Zhang Y., Nelson M., Nietfeldt J.W., Burbank D.E., van Etten J.L.;
 RT "Characterization of Chlorella virus PBCV-1 CviA11 restriction and
 RT modification system."
 RL Nucleic Acids Res. 20:5351-5356(1992).
 RN (2)
 RP REVISIONS, AND SEQUENCE FROM N.A.
 RX MEDLINE=95133167; PubMed=7831789;
 RA Lu Z., Li Y., Zhang Y., Kutish G.F., Rock D.L., van Etten J.L.;
 RT "Analysis of 45 kb of DNA located at the left end of the Chlorella
 RT virus PBCV-1 genome."
 RL Virology 206:339-352(1995).

CC -1 FUNCTION: THIS METHYLASE RECOGNIZES THE DOUBLE-STRANDED SEQUENCE
 CC CATC. CAUSES SPECIFIC METHYLATION ON A-2 ON BOTH STRANDS AND
 CC PROTECTS THE DNA FROM CLEAVAGE BY THE CVAII ENDONUCLEASE.
 CC -1 CATALYTIC ACTIVITY: S-adenosyl-L-methionine + DNA adenine = S-
 CC adenosyl-L-homocysteine + DNA 6-methylaminopurine.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL; M86634; AAC92381.1; .
 CC EMBL; U42580; AAC96619.1; .
 CC PIR; S27901; S27901.
 CC PIR; S35442; S35442.
 CC RBASE; 1542; M.CV:AIL.
 CC InterPro; IPR002294; D12N6_mtfase.
 CC InterPro; IPR002052; N6_Mtase.
 CC Pfam; PFC2086; MethyltransferaseD12; 1.
 CC PRINTS; PR03505; D12N6MTFRASE.
 CC PROSITE; PS00092; N6_MTFASE; 1.
 CC Transferase; Methyltransferase; Restriction system.
 CC KW
 CC SEQUENCE 326 AA; 37492 MW; 8CC1FC75D83F0D99 CRC64;

Query Match 27.7%; Score 48; DB 1; Length 326;
 Best Local Similarity 45.5%; Pred. No. 30;
 Matches 10; Conservative 2; Mismatches 10; Indels 0; Gaps 0;

QY 14 TIDLQAGSALPSAVALTSP 35
 I : : : : :
 DB 83 TIDLQAGSALPSAVALTSP 104

RESULT 25

RECA AZOVI
 ID RECA AZOVI STANDARD; PRT; 349 AA.
 AC P29246;
 DT 01 DEC 1992 (rel. 24, Created)
 DT 01-DEC-1992 (rel. 24, Last sequence update)
 DT 16-OCT-2001 (rel. 40, Last annotation update)
 DE RECA protein (Recombinase A).
 CN RECA.
 OS Azotobacter vinelandii.
 OC Bacteria; Proteobacteria; gamma subdivision; Pseudomonadaceae;
 OC Azotobacter.
 OX NCBI_TaxID=54;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=9225147; PubMed=1563632;
 RA Venkatesh T.V.; Das H.K.;
 RT "The Azotobacter vinelandii recA gene: sequence analysis and
 RT regulation of expression."
 RL Gene 113:47-53(1992).
 CC ! FUNCTION: CAN CATALYZE THE HYDROLYSIS OF ATP IN THE PRESENCE OF
 CC SINGLE STRANDED DNA, THE ATP-DEPENDENT UPTAKE OF SINGLE-STRANDED
 CC DNA BY DUPLEX DNA, AND THE ATP-DEPENDENT HYBRIDIZATION OF
 CC HOMOLOGOUS SINGLE-STRANDED DNAs. IT INTERACTS WITH LEXA CAUSING
 CC ITS ACTIVATION AND LEADING TO ITS AUTOCATALYTIC CLEAVAGE.
 CC -1 SUBCELLULAR LOCATION: Cytoplasmic (by similarity).
 CC -1 SIMILARITY: BELONGS TO THE RECA FAMILY.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----

CC EMBL; S96498; -; NOT_ANNOTATED_CDS.

DR PIR; JN0321; JN0321.
 DR HSSP; P03017; 2REB.
 DR InterPro; IPR003593; AAA_ATPase.
 DR InterPro; IPR001553; RecA.
 DR Pfam; PF00154; recA; 1.
 DR PRINTS; PR00142; RECA.
 DR ProDom; PD000229; RecA; 1.
 DR SMART; SM00382; AAA; 1.
 DR PROSITE; PS00321; RECA_1; 1.
 DR PROSITE; PS50162; RECA_2; 1.
 DR PROSITE; PS50163; RECA_3; 1.
 DR DNA damage; DNA recombination; SOS response; ATP-binding; DNA binding.
 FT NP BIND 65 72 ATP (BY SIMILARITY).
 SQ SEQUENCE 349 AA; 37203 MW; 9B1B2JD0EAEJAA29 CRC64;
 Query Match 27.7%; Score 48; DB 1; Length 349;
 Best Local Similarity 43.8%; Pred. No. 33;
 Matches 14; Conservative 6; Mismatches 10; Indels 2; Gaps 1;
 QY 1 VEKNITVTASVDPTI--DLLQAGSALPSAVA 30
 : : : : :
 DB 308 LEDNPEVAAAVEKSIRDQLLAAPASARPAALA 339
 Search completed: January 3, 2003, 13:01:58
 Job time : 15.3043 secs